Instruction

- You will evaluate the quality of the summaries of the six reports.
- Each report has 4 summaries. Please rate these four summaries according to the five criteria.
- Rate each summary on a scale from 1(worst) to 5(best) by Relevance, Consistency, Non-redundancy, Fluency, Coherence.

Definition

- Relevance: The rating measures whether the summary captures the key point of the article. Consider whether some or all of the crucial information is included in the summary.
- Consistency: The rating measures whether each sentence is well-placed for each information. Consider whether the sentence flows well.
- Non-redundancy: The rating measures whether the summary contains unnecessary repetition sentences. Redundant sentences do not mean only sentences with the same word composition. Where words have different uses but have the same meaning, we define them as duplicate sentences. If you think there are many redundancies, please indicate closer to 1 (worst).
- Fluency: The rating measures whether the summary is easy to read and understand. Consider the quality of the summary as a whole.
- Coherence: The rating measures whether the summary is well structured and well organized as a whole.

Report number

- The order list is based on "https://github.com/erinszeto/csci544-project/tree/main/data/abstraction/test"

1. Hee Ji - 95-408.txt ~ AIMD-96-54.txt
2. Erin - AIMD-96-82.txt ~GAO-01-46.txt
3. Danielle - GAO-01-493.txt~ GAO-01-781.txt
4. Surya - GAO-01-842.txt ~ GAO-02-398.txt
5. Haley - GAO-02-406.txt ~ GAO-02-634.txt

Find reports using this link:
1. labeled_extractive_summaries : https://github.com/erinszeto/csci544-project/tree/main/data/extractive-summaries/test
2. baseline_summaries : https://github.com/erinszeto/csci544-project/tree/main/data/baseline-summaries/test
3. attentive context_ summaries : https://github.com/erinszeto/csci544-project/tree/main/data/model-summaries/attentive-context-summaries/test-summaries
4. abstractive_summaries : https://github.com/erinszeto/csci544-project/tree/main/data/abstraction/test

Labeled extractive summaries

Relevance.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Consistency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Non-redundancy

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Fluency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Coherence

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Baseline summaries

Relevance.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Consistency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Non-redundancy

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Fluency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Coherence

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Attentive context summaries

Relevance.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Consistency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Non-redundancy

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Fluency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Coherence

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Abstractive summaries

Relevance.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Consistency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Non-redundancy

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Fluency.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best

Coherence

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

worst best